

SEQUENCE LISTING

109



(1) GENERAL INFORMATION:

(i) APPLICANT: Ullrich, Axel
 Aoki, Naohito
 Kim, Yeong Woong
 Wang, Hong Yang
 Chen, Zhengjun
 Naylor, Oliver
 Kharitonenkova, Alexei Igorevich

(ii) TITLE OF INVENTION: NOVEL PTP20, PCP-2, BDP1, CLK, AND SIRP POLYPEPTIDES AND RELATED PRODUCTS AND METHODS

(iii) NUMBER OF SEQUENCES: 38

(iv) CORRESPONDENCE ADDRESS:
 (A) ADDRESSEE: Lyon & Lyon
 (B) STREET: 633 West Fifth Street
 (C) CITY: Suite 4700
 (D) STATE: Los Angeles
 (E) COUNTRY: California
 (F) ZIP: U.S.A.
 90071-2066

(v) COMPUTER READABLE FORM:
 (A) MEDIUM TYPE: 3.5" Diskette, 1.44 Mb storage
 (B) COMPUTER: IBM Compatible
 (C) OPERATING SYSTEM: IBM P.C. DOS 5.0
 (D) SOFTWARE: FastSEQ for Windows 2.0

(vi) CURRENT APPLICATION DATA:
 (A) APPLICATION NUMBER: 08/877,150
 (B) FILING DATE: June 17, 1997
 (C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
 (A) APPLICATION NUMBER: U.S. 60/019,629
 (B) FILING DATE: June 17, 1996
 (A) APPLICATION NUMBER: U.S. 60/023,485
 (B) FILING DATE: August 9, 1996
 (A) APPLICATION NUMBER: U.S. 60/030,860
 (B) FILING DATE: November 13, 1996

(A) APPLICATION NUMBER: U.S. 60/034,286
(B) FILING DATE: December 19, 1996

(A) APPLICATION NUMBER: U.S. 60/030,964
(B) FILING DATE: November 15, 1996

(vii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Warburg, Richard J.
(B) REGISTRATION NUMBER: 32,327
(C) REFERENCE/DOCKET NUMBER: 225/298

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: (213) 489-1600
(B) TELEFAX: (213) 955-0440
(C) TELEX: 67-3510

(2) INFORMATION FOR SEQ ID NO: 1:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 6 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in positions 3 and 5 stands for an unspecified amino acid.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

Phe Trp Xaa Met Xaa Trp
1 5

(2) INFORMATION FOR SEQ ID NO: 2:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in position 6 stands for either Ser, Ile or Val.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

His Cys Ser Ala Gly Xaa Gly
1 5

(2) INFORMATION FOR SEQ ID NO: 3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:

Phe Leu Glu Arg Leu Glu
1 5

(2) INFORMATION FOR SEQ ID NO: 4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in positions 3 and 5 stands
for an unspecified amino acid.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:

Arg Trp Xaa Met Xaa Trp
1 5

(2) INFORMATION FOR SEQ ID NO: 5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in position 6 stands for either Ser, Ile or Val.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

His Cys Ser Ala Gly Xaa Gly
1 5

(2) INFORMATION FOR SEQ ID NO: 6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 27 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:

CTCTGTGTCC ACAGCAGTGC TGGCTGT

27

(2) INFORMATION FOR SEQ ID NO: 7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 7 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

His Arg Asp Leu Ala Ala Arg
1 5

(2) INFORMATION FOR SEQ ID NO: 8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(ix) FEATURE:

(D) OTHER INFORMATION: "Xaa" in position 2 stands for Val or Met. "Xaa" in position 5 stands for Tyr or Phe.

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:

Asp Xaa Trp Ser Xaa Gly
1 5

(2) INFORMATION FOR SEQ ID NO: 9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 9:

CGGGATCCCT TCGCCTTGCA GCTTTGTC

28

(2) INFORMATION FOR SEQ ID NO: 10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 10:

CGGAATTCCCT AGACTGATAC AGTCTGTAAG

30

(2) INFORMATION FOR SEQ ID NO: 11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 11:

Asp Leu Lys Pro Glu Asn
1 5

(2) INFORMATION FOR SEQ ID NO: 12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 12:
 Ala Met Met Glu Arg Ile
 1 5

(2) INFORMATION FOR SEQ ID NO: 13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 13:

TATAGCGGCC GCTAGACTGA TACAGTCTGT

30

(2) INFORMATION FOR SEQ ID NO: 14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 32 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 14:

TCCCCCGGGGA TGCCCCATCC CCGAAGGTAC CA

32

(2) INFORMATION FOR SEQ ID NO: 15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 15:

TATAGCGGCC GCTCACCGAC TGATATCCCG ACTGGAGTC

39

(2) INFORMATION FOR SEQ ID NO: 16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 16:
TCCCCCGGGG AGACGATGCA TCACTGTAAG

30

(2) INFORMATION FOR SEQ ID NO: 17:
(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 17:
TATAGCGGCC GCGCTGGCCT GCACCTGTCA TCTGCTGGG

39

(2) INFORMATION FOR SEQ ID NO: 18:
(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 18:
CGGAATTCCAT CGGGCATTCC AACCGAACTC

30

(2) INFORMATION FOR SEQ ID NO: 19:
(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 39 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 19:
TATAGCGGCC GCCCTGACTC CCACTCATT CTTTTTAA

39

(2) INFORMATION FOR SEQ ID NO: 20:
(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 20:
 CGGAATTCCG CCACCATGGC CCCTATCTA GGTTAT

36

(2) INFORMATION FOR SEQ ID NO: 21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 36 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 21:
 GCCAAGCTTG CCACCATGGC CCCTATCTA GGTTAT

36

(2) INFORMATION FOR SEQ ID NO: 22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 22:
 GTAGCAGTAA GAATAGTTAA A

21

(2) INFORMATION FOR SEQ ID NO: 23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 23:
 GTTGCCTGA GGATCATTA A GAAT

24

(2) INFORMATION FOR SEQ ID NO: 24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 24:
 GTTGCCTCTGA GGATCATCCG GAAT

24

(2) INFORMATION FOR SEQ ID NO: 25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 25:
 TACAATTCTC ACTGCTACAT GTAAGCCATC

30

(2) INFORMATION FOR SEQ ID NO: 26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 26:

Pro Ile Tyr Ser Phe Ile Gly Gly Glu His Phe Pro Arg
 1 5 10

(2) INFORMATION FOR SEQ ID NO: 27:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 9 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 27:

Ile Val Glu Pro Asp Thr Glu Ile Lys
 1 5

(2) INFORMATION FOR SEQ ID NO: 28:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 6 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 28:

Tyr Gly Phe Ser Pro Arg
1 5

(2) INFORMATION FOR SEQ ID NO: 29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 12 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 29:

Ile Lys Glu Val Ala His Val Asn Leu Glu Val Arg
1 5 10

(2) INFORMATION FOR SEQ ID NO: 30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 8 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 30:

Val Ala Ala Gly Asp Ser Ala Thr
1 5

(2) INFORMATION FOR SEQ ID NO: 31:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2226 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 31:

GAATTCCGGC	ACGAGGGCGG	TTGCAGTATG	AGTCGCCAAT	CGGACCTAGT	GAGGAGCTTC	50
TTGGAGCAGC	AGGAGGCCG	GGACCCCG	AAAGGGCCAA	TCCCTGGCCG	TGAGGTCAGC	120
GACATTAAAG	CCCGCTCACT	GGCTTGGAG	ACTGAACTGTG	TGTGCTCCAC	AAAGGCCGC	180
ACTCACCGC	AAAACCTAA	GAACACCCG	TACAAAAGAC	TGTTACCGTA	TGATGAGACG	240
AGAGTCATCC	TTTCCTCGCT	CCAGGAGGAA	GGACACGGAG	ATTACATTTA	TGCCAACITC	300
ATCCGGGGCA	CAGATGGAG	CCAGGCCCTAC	ATTCGGACGC	AAGGACCCCT	GCTTCACACT	360
CTGTTGACT	TCTGGCCCT	GOTTTGGAG	TTTGGAACTCA	AGGTGATCTT	GATGCCCTGT	420
CAGGAGACAG	AAAATGGACG	GAGGAAGTGT	GAACAGCTACT	GGGGCCAGGA	GCGGGAGGCT	480
CTACAGGCCG	GGCCTTCTG	CATCACCCCT	AAACAGAGAA	GAGCACTGAC	TTCGGACATC	540
ACTCTCAGGA	CCCTCCAGGT	TACATTTCAAC	AAGGAATCCC	GTCTCTGTGCA	CCAGCTACAG	600
TACATGTTCT	GGGGCGACCA	GGGGTTCTCA	AGCAGTTCCG	ATCACATCTT	CACCATGGTG	660
GAGGGAGGCG	GGTGGCTCCA	AGGACTTGG	CTTGGGACCC	TCTCTGTCGA	CTGCGATGCT	720
GCTCTGGAC	GAACAGCTGT	CTTGTGTCG	GTGATTACAG	TGAGGCACTT	GCTTCAGACT	780
CAGAACATTC	CACCCAAATT	CAGCCCTTTT	GAAGGTGTC	TGGAGATGCG	AAAGAACGGCA	840
CTCTGAGCGG	TGCAAGACA	GGAGCGACTA	AGGTTCTCTG	ACACACAGAT	GCTCTAGCTA	900
TTCTCCCGCA	CTCTCCAGAA	CAACAGCTCC	CTCTACAGCA	ACCTCAAGGA	GAACCGCGCT	960
CCAATCTGCA	AAGCTCTGC	CTTCCCCTAGG	ACCTCTCAG	CCCTGCTGCG	CACATCCCGC	1020
CCACTGGGTG	GGCTTCTCG	GAGCATCTCG	GTGCTCTGGC	CACCGGACCT	TCCCATGCGG	1080
GACACTTACG	CTGTGGTGC	GAAGCTGTC	GTCTCCGGC	GCACAGGGCC	GGCACCGGG	1140
GGGCCAACAA	GCACGACAC	CCCGATCTAC	AGCAGCTGG	CTCCACCTAT	CCAGCGGCC	1200
GTGTCAACACA	CGAAAAGCCG	GCAGGGGCA	ACCGACTCG	GCCGAGTTCC	TGCGGATGAA	1260
AACCTTCCG	GGCCTGATGC	CTATGAGGAA	GTAAACAGAT	GAGGCCAGAC	TGTTGGGCTA	1320
GCTTCAACT	GGGGCTATGG	AAGACCTAAA	GGGGCACCGG	ATCTCTCAGC	GGAGTGGACAA	1380
CGGGTGAAT	GAGTGTGTA	CCAGTTCCAG	CTCTGCTACTC	AGTGTGTCG	GGGCACCTGC	1440
AACCCCATC	CTGCTGTG	CTGTCTTATG	TATGAGTGG	ACTCATGGG	CTGAATCAA	1500
ATAAAAGTT	CTCAGGGTAG	AAAAAAACAA	ATAGGACTCT	TGGGCACTGG	TTATAGCACT	1560
CAAACCGAG	GGCTAGGAGG	GGTAAGTGGG	GGAGGTGGT	GATCTACTCT	GAGAAAGTTT	1620
AGGAAGAAC	ATCAAGGCT	AGCATGCCA	CTCTCTCCC	CATACACCTA	CTGGGAACTG	1680
CACCCAGAC	AGAGCTCTAA	CTTGAGCATG	CAACCTCGAC	AGGTGCTAC	CTGGATGGAC	1740
ATGCTGGCC	TACAGCTTAG	GACATGCTTA	ATTAGATCTT	CTGTTAAACT	TGCAATGAGC	1800
TAGAAAGATC	TCCGCTCTGT	CAGGGAAATG	GATCACCTAG	TCAGGTAAAT	AGTGTGCGAT	1860
CCAGAGACAA	GAACCTCAAG	ATACAGCTTT	TCTCAAATTA	GAAGAAAAATA	GATCTCAAG	1920
AATAAAATGA	TGTCATCAATG	TCTACGCCCT	GATCTCTGGC	TGCTCTACTG	CCATAATGTC	1980
ACAAACAAAGT	CAGGCTCTAT	ATGACACTTG	TTTCATCTAGT	CAGTGGTGC	TGTGGCTCT	2040
GCAGGCTCAG	ATAGTGCCTT	CTGCAGACTC	TTGGAATGCC	CGTCTTGAAAC	TTGATGAAAG	2100
CTTCTACCGG	GAACCTGTA	ACATCATTAA	AATTATTAAT	GTAGAATTCA	ATAAAAGACTG	2160
GGTCAAAAC	TCNAAAAAAA	AAAAAAA	AAAAAAAC	TCGAGAGTAC	TTCCTAGAGCG	2220
	GGCGGG					2226

(2) INFORMATION FOR SEQ ID NO: 32:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 453 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 32:

Met	Ser	Arg	Gln	Ser	Asp	Leu	Val	Arg	Ser	Phe	Leu	Glu	Gln	Gln	Glu
1				5				10				15			

Ala	Arg	Asp	His	Arg	Lys	Gly	Ala	Ile	Leu	Ala	Arg	Glu	Phe	Ser	Asp
20				25							30				

Ile Lys Ala Arg Ser Val Ala Trp Lys Thr Glu Gly Val Cys Ser Thr
 35 40 45
 Lys Ala Gly Ser Gln Gln Gly Asn Ser Lys Lys Asn Arg Tyr Lys Asp
 50 55 60
 Val Val Pro Tyr Asp Glu Thr Arg Val Ile Leu Ser Leu Leu Gln Glu
 65 70 75 80
 Glu Gly His Gly Asp Tyr Ile Asn Ala Asn Phe Ile Arg Gly Thr Asp
 85 90 95
 Gly Ser Gln Ala Tyr Ile Ala Thr Gln Gly Pro Leu Pro His Thr Leu
 100 105 110
 Leu Asp Phe Trp Arg Leu Val Trp Glu Phe Gly Ile Lys Val Ile Leu
 115 120 125
 Met Ala Cys Gln Glu Thr Glu Asn Gly Arg Arg Lys Cys Glu Arg Tyr
 130 135 140
 Trp Ala Gln Glu Arg Glu Pro Leu Gln Ala Gly Pro Phe Cys Ile Thr
 145 150 155 160
 Leu Thr Lys Glu Thr Ala Leu Thr Ser Asp Ile Thr Leu Arg Thr Leu
 165 170 175
 Gln Val Thr Phe Gln Lys Glu Ser Arg Pro Val His Gln Leu Gln Tyr
 180 185 190
 Met Ser Trp Pro Asp His Gly Val Pro Ser Ser Asp His Ile Leu
 195 200 205
 Thr Met Val Glu Glu Ala Arg Cys Leu Gln Gly Leu Gly Pro Gly Pro
 210 215 220
 Leu Cys Val His Cys Ser Ala Gly Cys Gly Arg Thr Gly Val Leu Cys
 225 230 235 240
 Ala Val Asp Tyr Val Arg Gln Leu Leu Leu Thr Gln Thr Ile Pro Pro
 245 250 255
 Asn Phe Ser Leu Phe Glu Val Val Leu Glu Met Arg Lys Gln Arg Pro
 260 265 270
 Ala Ala Val Gln Thr Glu Glu Gln Tyr Arg Phe Leu Tyr His Thr Val
 275 280 285
 Ala Gln Leu Phe Ser Arg Thr Leu Gln Asn Asn Ser Pro Leu Tyr Gln
 290 295 300
 Asn Leu Lys Glu Asn Arg Ala Pro Ile Cys Lys Asp Ser Ser Ser Leu
 305 310 315 320
 Arg Thr Ser Ser Ala Leu Pro Ala Thr Ser Arg Pro Leu Gly Gly Val
 325 330 335
 Leu Arg Ser Ile Ser Val Pro Gly Pro Pro Thr Leu Pro Met Ala Asp
 340 345 350

Thr Tyr Ala Val Val Gln Lys Arg Gly Ala Ser Gly Ser Thr Gly Pro
 355 360 365
 Gly Thr Arg Ala Pro Asn Ser Thr Asp Thr Pro Ile Tyr Ser Gln Val
 370 375 380
 Ala Pro Arg Ile Gln Arg Pro Val Ser His Thr Glu Asn Ala Gln Gly
 385 390 395 400
 Thr Thr Ala Leu Gly Arg Val Pro Ala Asp Glu Asn Pro Ser Gly Pro
 405 410 415
 Asp Ala Tyr Glu Glu Val Thr Asp Gly Ala Gln Thr Gly Gly Leu Gly
 420 425 430
 Phe Asn Leu Arg Ile Gly Arg Pro Lys Gly Pro Arg Asp Pro Pro Ala
 435 440 445
 Glu Trp Thr Arg Val
 450

(2) INFORMATION FOR SEQ ID NO: 33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5581 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 33:

AATTCCGGGC GCCAGTCCCG CTCCCGGCCG CGCCGCTCCG CTCCGGCTCG GGCTCCGGCT
 CGCCCTGGGG TGGGCTGGGG CTCCGGGGGG CGCGCTCCCG CGGGAGCGCC
 CGACCTCCAA CCATGGGGGG CGGCGGGGGG CTCGCTCG CACTCACCTTCC CCAAGCTCTGC
 GCGCGGGAGA CGGAGACTCC GGCAGCTGGC TGACCGCTTC AGGAGGGCAAG TGACCCAGCA
 GTGCGCTGGG AGTACAGCCA CGCCGAGCTTG GATGACTCTTG AGTGGGAGCA AGTGGGAATC
 CACCCATGGGG CGCCGAGCTTG CGCCGAGCTTG CCCCAAGGCT CCTACTTGAT GGTCAAACACT
 TCCCAAGCTG CGCCGAGCTTG CGCCGAGCTTG AGAGCCCTGAG CGAGAAATGAT
 ACCCACTGTG TGCAAGCTGAG CTACATCTCTG TACAGGGGGG ACGGCACAGG CGGGCACCTG
 CGCGCTTACG TCCGGCTTAA TGGGGGGGG CGTGGCGAGTG CTGTTGGAATGATGAGTGG
 TCCCAAGGGC GTCAAGTGGCA CGAGGCTGAG CTGGCTGTCA GCACCTTCTG GCCCAATGAA
 TATCAGGTG TGTTGAGGCG CCTCATCTCC CGAGAGCGCA GGGGCTACAT GGGCTCTAGAT
 GACATCTGCA TTCTCACTCA CGCCCTGGCGA AAAGGCCCGA ATCTCTCCG CTCGGGCAC
 GTGGAGGCTCA CGCCGAGGCCA GAACCGCTGG TTCCAGTGC TGCCAGCTGG AGAGCCCATG
 CGCCCAAGCT TCCCTCTGCA CGCCGAGAGC GGGGCGCTGG TGCCGGCGGG CGCCGTGCGC
 ACATCAGCCA CGCGCTTCTCTG CGCCACTTCC CGCGCTGGCTG CGTGGAGCC CGCCGAGCAG
 GACCTGTACG GTCTGTGTTGCG CAAGGGCCCC CGGGGGGGGG TCTCTAACCTT CGCGAGCTC
 ATCGTCAAGG AGCCCCAAC TCCCATCGCC CCCCCACAGC TGCTGGCTGC TGCCCCCAC
 TACCTCATCA TCCAGCTCAA CACCAACTCC ATCATGGGG ACGGGCGGAT CGTGGCAAG
 GAGATTTGAGT ACCGGCATGGC GCGCGGGCGG TGGGCTGGG TGACCGCCGT CAGCCTGCA
 ACCTACAAAGC TGTTGGACCT CGACCCCGAC ACAGACTATC AGATCAGCTG GCTGCTCACG
 CGTCCCGGAG ACGGCGGCAAC TGCCGCTGGG GCGCAACCTTCA ATCAGCGCGA CAAAATGGG
 AGAGCCCATG AGGGCCCCAA AGGCGCTGGCT TTGCTGTGAGA TCCAGGCCG TCAGCTGAC
 CTGGAGTGGG AACCCACTGGG CTACAAAGCTG AGCGCTTGGC ACACCTATAC TGTGTCGCTG
 CTGCTACTC ACACCTCTGG CGCGAGCCAC AACCAGACCA TCCAGAGAGTG TGTGAAGACAA
 GAGCAAGGTG TCAGCCGCTA CACCATCAAG AACCTGTGCG CCTATCGGAA CGTTCACGTG
 AGGCTTGTC TCACTAAACCC TGAGGGGGCG AAGAGGGCGA AGGAGGGTCAC TTTCCAGACG
 GATGAGGATG TGCCCACTGG GATTCAGGCC CGCCCTGAGA CCTTCACCTCC ACTGGAGGAC

ATGATCTTCC	TCAAGTGGGA	GGAGCCCCAG	GAGCCCCAATG	GTCTCATCAC	CCAGTATGAG	1680
ATCAGCTACC	AGAGCATCGA	GTCATCACAG	CCGGCAGTGA	ACGTGCGGAGG	CCCCACGACT	1740
ACCATCTTCA	AGCTCCCAA	TGAGACCTAC	ATGTCCTCT	CCAAACCTGCA	CCCCAGGCAC	1800
ACCTACCTGT	TCTCGGTGCG	GGCCGCCACA	GGCAAAAGGT	TCGGCCAGGC	GGCACTCCT	1860
GAGATTAACCA	CTAAATCTC	TGCTCCCAAG	TTGATTTATG	CCGACATGCC	GTCACCCCTG	1920
GGGGGACTCTG	AAACACCAT	CAGCTTGCTG	CTGAGGGCCG	CAACAGGCCG	CGGTGCCCC	1980
ATCAGTGTGT	ACCAAGTGAT	TGTGGAGGAG	GAGCCGGCGC	GAGGCTCGG	CGGGAGCAGG	2040
TGGACAGGAC	TGCTTCCAG	TGCCATTGAC	CTTCGAGGGC	CGCCTGGCCC	CAGGCTGGTG	2100
CACTACTTCG	GGGCGGAACT	GGCGCCAGC	AGTCTACCTG	AGGCCATGCC	CTTACCGTG	2160
GGTGACAACC	AGAACCTACCG	AGGCTTCTCG	AACCCACAC	TTGAGCCTAG	GAAGGGCTAT	2220
CTCATCTACT	TCCAGGAGC	AAAGCCACCT	AAAGGGGAGA	CCCGCTGAA	TTGATCCC	2280
ATIGGAGGAG	AAAGCTCTCG	CAAGGAAAGC	AAAGGGCCCC	TGGAGGTGTC	CCAGAGATCG	2340
GAGGAGATGA	GGGCTTATCT	GGGCATCTG	CGACGGGGGG	TCGCTGCTC	CATCTCTTC	2400
CTGGTGGCCA	TCATCTGCT	CATCCGCAA	GGGAAGCCG	TAACATGAC	CAAGGCCAAC	2460
GTCAACTACC	GCCAGAGAAA	GACACATA	ATACGGCCG	TGGACCCAG	CTTCACAGAC	2520
CAAGGACACC	TGCGAGGAGA	CGAGCACATA	GGGCTCTCT	TCATGGACAC	CCATGACACT	2580
ACCCACCGGG	GGAGACAGCG	CGACCGTGG	GTCACTGAGG	CCACGACCT	CTTGGGGGGC	2640
TECCCGAGGG	GTCTCCCTG	CGGAGGGC	TCCCCATACC	ACAGGGGGCA	GCTGACACCT	2700
GGCGTGTGCG	TCCGACGACT	TGTCGACCA	ATCAACAGCA	TAAGAACGGC	CGAGGGTTAC	2760
GCGCTTAACG	AGGAGTATG	GAGCTTCTTT	GAAGGGTGG	ACGCCACAAA	GAAGAAAGAC	2820
AAAGGTAAAG	CGACCGCGA	GGGACCAATG	CTCTGCTTAT	ATCGCACCG	ACTGAACTG	2880
CAACCGATTC	TGGGGAGACCC	CAATGCCGAC	TACAATTAAG	CCAAACTACAT	AGA-TGGTTAC	2940
CAACGGTCAA	ACCACTTAC	AGGCCATCAA	GGGGCGAAGG	CTGAGATGGT	CTATGAACTTC	3000
TGGCGTATGG	TGTGAGGAG	GAACCTGTTG	AGCATCGTC	TGATCACCAC	GCTGTTGAG	3060
GTGGCGAGG	TGAATGTC	AGCGTACTGG	CGGCGAGACT	CAGACACCTA	CGGGGACATC	3120
AAGATTATGC	TGGTGAAGG	AGAACGACCC	GTCTGAGTAT	CTGTTGGCAC	TTTTGACCTG	3180
GAGCGGAGAG	GCTACTCTG	CGGGCACAGG	GTGCGGCACT	CCACACTTCAC	AGCGTGGGCC	3240
GAGCATGGC	TCCCCCTAAC	TGCCCCGGG	CTGCTGGCT	TCATCCCGGG	GOTGAAAGCC	3300
TCCACCCAC	CTGATCGCGG	ATCTACATG	GGCGGGGAC	GGCCGCCCCA	CGGCCGACAA	3360
COTTOCTATA	TGCTCTCTG	TGTGATGCT	GACATGGCA	AGTGTGAGGG	COTCTGTGAC	3420
ATTACAACAT	GTGTGAGAAC	TGCTCTGCTC	CGGCGTGTCA	ACATGATTC	GACTGAGGG	3480
CACTACATCT	TCTTACATGA	TGCACTTCT	GGGGCTCTC	TGTGTGGGA	GACCAACATC	3540
CTCTGCTACTG	AGTCTAACGG	CACCTAACAG	GGAGATGAT	GCATTGATTC	TCAGAGTAAT	3600
TCTCTCCAGG	TGGGGAGA	CTTCCAGACG	CTGAACTCG	TCACCCGGG	GCTGAGCTG	3660
GAGGAGTGTCA	GCATCCCTCT	GTGCGGCGG	AACCGCGACA	AAACCGGCGA	CATGGGACGT	3720
TCCGGCCCG	ACCGCTGGC	GGCTCTTCTC	ATCTCCACTG	ATGGGGACTC	CAACAACTAC	3780
ATTAGTACG	CCTCTGAGTA	CGACTGCTG	GGCAAGGCTG	CCTCTCATGGT	GACCCCTGCA	3840
CGCGCTGAGG	GGCACCCAGC	GGACTCTTCA	CGGGCTGGCT	ACGATTAACGG	GTGACCTCC	3900
ATCGCTATCG	TCACCAAGCT	GAACCCATTC	AACCTCCGCT	GGCCCTTGCT	CGACTACTGG	3960
CCAGAGGCG	GGGGCGGAA	ATATGGCTC	ATGGAGGTGA	AGTTTATGTC	GGGCACAGCT	4020
GATCGAAAGCT	TAGTGGCTG	AGTCTTCGG	GTGCGAACAA	TCTCTCGGGT	GCAGGAGGGA	4080
GACCTGCTG	TGGGGAGACTT	CGCTGTTCTG	CGCTGTGCTG	ATACAGGGGA	CACACCTGCA	4140
TCCAAGAAGG	CCTCTCTGCA	CTCTGCTGGT	GAGGTGGACG	AGTGGCAGGC	CGAGAGTGGG	4200
GATGGCGCA	CCATCTGTC	CTGCTAACAC	GGGGGGAGA	CGACGCGCAC	CTTCTGGGCC	4260
TGCGCCAGG	TCTCTGAGAT	GATCGCTGTC	CACAACTTG	TGACCTTTT	TTTGTGTC	4320
CAAACCTCG	GGAAACTACAA	ACCCAACTTA	GTGAGACCCA	TGATGACTGA	CCACCTTTTG	4380
TACGATGTTG	TCTCTGAGAT	CTTGGAGGGG	TGGACTGAG	TAAGCGGGG	CCCTGGGCTG	4440
GGGGCACCCAC	GGGGCGAAC	CACCATCTCTG	GACTGGCGAG	GAAGATCAGT	4500	
GCCTCTCTG	CTGCCCCAAC	ACACTCCCA	GGGGCAAGCA	TCGACTGGA	TGCTGGGCTA	4560
TCTCTGCTCC	CTTTCACACTG	TGGGGAGGG	CTTTCGCTTG	TCCCAATGGC	GGGTGGTGGG	4620
CCAAGGAGGA	GCTTGTAGCA	TGTCACCCAC	ACCCCCACCT	CCATAGGGT	CTGAGGCC	4680
GTGCTGAGAG	GGCTGGTGTG	GGCTGGGAGA	GTGACAAGG	CTCAAGACCG	CTGGCTCTGG	4740
GGGACTCTGG	CCAAAGGGGGT	TGGCAGGATC	CTGGGTTTG	GGAGGGATGA	GTAGGGCCCT	4800
GCAGAGGAGC	TCCAGGAGCA	AGGTCTCCAC	TCAGGCTGCG	CCCTCTGCA	GTGGGTAGAG	4860
GATGTAATG	GACTTGGCAT	TTAGGATTC	ATCTGGGGCA	CCCCCTGAAG	GTCCCCCCCC	4920
AGCAAGTCTC	AATTCTGATA	GGCACTGGGG	CACACTGACT	GTCTCTCCCA	GGGAAACTGC	4980
ACCCCTCTCC	TCCCCACTG	CCCCCTGAGC	CCCTGAGATA	TTTTGCTCAC	TATCCCTCCC	5040
CACTGGCTTC	CTGTGATGAT	GGCTCTGACT	CCCTGAAACCA	GGATCTGCT	ATTACTGCTG	5100
TCCCATGGGG	GGCTCTCTTC	CTGCCATGAC	CACTGFIGCA	GAATGAAGTC	ACCTCGCCCC	5160
CCTCTCTCTT	TAATCTCTAG	GGCTCTACTG	CTGTGCTTGC	TCAGCTTGGG	CCAGTGACAA	5220
TCTGCAAGG	TGAACACAG	CCCCCTGGGT	TGAGGGCCCT	GTGAGCTCTG	GTAGGGCTGC	5280
CCGGTGTGGG	GAGGGGGAGT	GTAGAGACAG	GGCTGTCAT	ACCCCTCTGGA	GTGAGAGCA	5340

AGAGGGTAGGA	CCAGTGCTTT	TTTGTGTTCTT	TTGTTTATTCTT	TGGTTGGGTG	GGTGGGAAGG	5400
TCTCTTAA	ATGGGGCAGG	CCACACCCCC	ATTCGGTGCC	TCAATTCCCC	CATCTGTA	5460
CTGTAGATAT	GACTACTGAC	CTACACTCGCA	GGGGGCGTGTG	GGGAGGCATA	AGCTGATGTT	5520
TGTAAAGCGC	TTTGTAAATA	AACGTGCTCT	CTGAATGCCA	AAAAAAA	AACRAAAA	5580
A						5581

(2) INFORMATION FOR SEQ ID NO: 34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1430 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 34:

Met Ala Arg Ala Gln Ala Leu Val Leu Ala Leu Thr Phe Gln Leu Cys						
1	5	10	15			
Ala Pro Glu Thr Glu Thr Pro Ala Ala Gly Cys Thr Phe Glu Glu Ala						
20	25	30				
Ser Asp Pro Ala Val Pro Cys Glu Tyr Ser Gln Ala Gln Tyr Asp Asp						
35	40	45				
Phe Gln Trp Glu Gln Val Arg Ile His Pro Gly Thr Arg Ala Pro Ala						
50	55	60				
Asp Leu Pro His Gly Ser Tyr Leu Met Val Asn Thr Ser Gln His Ala						
65	70	75	80			
Pro Gly Gln Arg Ala His Val Ile Phe Gln Ser Leu Ser Glu Asn Asp						
85	90	95				
Thr His Cys Val Gln Phe Ser Tyr Phe Leu Tyr Ser Arg Asp Gly Thr						
100	105	110				
Gly Gly Thr Leu Arg Val Tyr Val Arg Val Asn Gly Gly Pro Leu Ala						
115	120	125				
Ser Ala Val Trp Asn Met Thr Gly Ser His Gly Arg Gln Trp His Gln						
130	135	140				
Ala Glu Leu Ala Val Ser Thr Phe Trp Pro Asn Glu Tyr Gln Val Leu						
145	150	155	160			
Phe Glu Ala Leu Ile Ser Pro Asp Arg Arg Gly Tyr Met Gly Leu Asp						
165	170	175				
Asp Ile Leu Leu Ser Tyr Pro Cys Ala Lys Ala Pro His Phe Ser						
180	185	190				
Arg Leu Gly Asp Val Glu Val Asn Ala Gly Gln Asn Ala Ser Phe Gln						
195	200	205				
Cys Met Ala Ala Gly Glu Pro Met Arg Gln Arg Phe Leu Leu Gln Arg						
210	215	220				

Gln Ser Gly Ala Leu Val Pro Ala Gly Ala Phe Gly Thr Ser Ala Thr
 225 230 235 240
 Gly Phe Leu Ala Thr Phe Pro Leu Ala Ala Val Ser Arg Ala Glu Gln
 245 250 255
 Asp Leu Tyr Arg Cys Val Ser Gln Ala Pro Arg Gly Gly Val Ser Asn
 260 265 270
 Phe Pro Glu Leu Ile Val Lys Glu Pro Pro Thr Pro Ile Ala Pro Pro
 275 280 285
 Gln Leu Leu Arg Ala Gly Pro Thr Tyr Leu Ile Ile Gln Leu Asn Thr
 290 295 300
 Asn Ser Ile Ile Gly Asp Gly Pro Ile Val Arg Lys Glu Ile Glu Tyr
 305 310 315 320
 Arg Met Ala Arg Gly Pro Trp Ala Glu Val His Ala Val Ser Leu Gln
 325 330 335
 Thr Tyr Lys Leu Trp His Leu Asp Pro Asp Thr Asp Tyr Glu Ile Ser
 340 345 350
 Val Leu Leu Thr Arg Pro Gly Asp Gly Gly Thr Gly Arg Trp Ala Thr
 355 360 365
 Pro His Gln Pro His Gln Met Arg Arg Ala His Glu Gly Pro Lys Gly
 370 375 380
 Leu Ala Phe Ala Glu Ile Gln Ala Arg Gln Leu Thr Leu Gln Trp Glu
 385 390 395 400
 Pro Leu Gly Tyr Asn Val Thr Arg Cys His Thr Tyr Thr Val Ser Leu
 405 410 415
 Cys Tyr His Tyr Thr Leu Gly Ser Ser His Asn Gln Thr Ile Arg Glu
 420 425 430
 Cys Val Lys Thr Glu Gln Gly Val Ser Arg Tyr Thr Ile Lys Asn Leu
 435 440 445
 Leu Pro Tyr Arg Asn Val His Val Arg Leu Val Thr Asn Pro Glu
 450 455 460
 Gly Arg Lys Glu Gly Lys Glu Val Thr Phe Gln Thr Asp Glu Asp Val
 465 470 475 480
 Pro Ser Gly Ile Ala Ala Glu Ser Leu Thr Phe Thr Pro Leu Glu Asp
 485 490 495
 Met Ile Phe Leu Lys Trp Glu Glu Pro Gln Glu Pro Asn Gly Leu Ile
 500 505 510
 Thr Gln Tyr Glu Ile Ser Tyr Gln Ser Ile Glu Ser Ser Asp Pro Ala
 515 520 525
 Val Asn Val Pro Gly Pro Arg Arg Thr Ile Ser Lys Leu Arg Asn Glu
 530 535 540

Thr Tyr His Val Phe Ser Asn Leu His Pro Gly Thr Thr Tyr Leu Phe
 545 550 555 560
 Ser Val Arg Ala Arg Thr Gly Lys Gly Phe Gly Gln Ala Ala Leu Thr
 565 570 575
 Glu Ile Thr Thr Asn Ile Ser Ala Pro Ser Phe Asp Tyr Ala Asp Met
 580 585 590
 Pro Ser Pro Leu Gly Glu Ser Glu Asn Thr Ile Thr Val Leu Leu Arg
 595 600 605
 Pro Ala Gln Gly Arg Gly Ala Pro Ile Ser Val Tyr Gln Val Ile Val
 610 615 620
 Glu Glu Glu Arg Ala Arg Gly Cys Gly Gly Thr Arg Trp Thr Gly Leu
 625 630 635 640
 Leu Pro Ser Ala Ile Asp Leu Arg Gly Gly Ala Gly Pro Arg Leu Val
 645 650 655
 His Tyr Phe Gly Ala Glu Leu Ala Ala Ser Ser Leu Pro Glu Ala Met
 660 665 670
 Pro Phe Thr Val Gly Asp Asn Gln Thr Tyr Arg Gly Phe Trp Asn Pro
 675 680 685
 Pro Leu Glu Pro Arg Lys Ala Tyr Leu Ile Tyr Phe Gln Ala Ala Ser
 690 695 700
 His Leu Lys Gly Glu Thr Arg Leu Asn Cys Ile Arg Ile Ala Arg Lys
 705 710 715 720
 Ala Ala Cys Lys Glu Ser Lys Arg Pro Leu Glu Val Ser Gln Arg Ser
 725 730 735
 Glu Glu Met Gly Leu Ile Leu Gly Ile Cys Ala Gly Gly Leu Ala Val
 740 745 750
 Leu Ile Leu Leu Leu Gly Ala Ile Ile Val Ile Ile Arg Lys Gly Lys
 755 760 765
 Pro Val Asn Met Thr Lys Ala Thr Val Asn Tyr Arg Gln Glu Lys Thr
 770 775 780
 His Met Ile Ser Ala Val Asp Arg Ser Phe Thr Asp Gln Ser Thr Leu
 785 790 795 800
 Gln Glu Asp Glu Arg Leu Gly Leu Ser Phe Met Asp Thr His Gly Tyr
 805 810 815
 Ser Thr Arg Gly Asp Gln Arg Ser Gly Gly Val Thr Glu Ala Ser Ser
 820 825 830
 Leu Leu Gly Gly Ser Pro Arg Arg Pro Cys Gly Arg Lys Gly Ser Pro
 835 840 845
 Tyr His Thr Gly Gln Leu His Pro Ala Val Arg Val Ala Asp Leu Leu
 850 855 860
 Gln His Ile Asn Gln Met Lys Thr Ala Glu Gly Tyr Gly Phe Lys Gln
 865 870 875 880

Glu Tyr Glu Ser Phe Phe Glu Gly Trp Asp Ala Thr Lys Lys Lys Asp
 885 890 895
 Lys Val Lys Gly Ser Arg Gln Glu Pro Met Pro Ala Tyr Asp Arg His
 900 905 910
 Arg Val Lys Leu His Pro Met Leu Gly Asp Pro Asn Ala Asp Tyr Ile
 915 920 925
 Asn Ala Asn Tyr Ile Asp Gly Tyr His Arg Ser Asn His Phe Ile Ala
 930 935 940
 Thr Gln Gly Pro Lys Pro Glu Met Val Tyr Asp Phe Trp Arg Met Val
 945 950 955 960
 Trp Gln Glu His Cys Ser Ser Ile Val Met Ile Thr Lys Leu Val Glu
 965 970 975
 Val Gly Arg Val Lys Cys Ser Arg Tyr Trp Pro Glu Asp Ser Asp Thr
 980 985 990
 Tyr Gly Asp Ile Lys Ile Met Leu Val Lys Thr Glu Thr Leu Ala Glu
 995 1000 1005
 Tyr Val Val Arg Thr Phe Ala Leu Glu Arg Arg Gly Tyr Ser Ala Arg
 1010 1015 1020
 His Glu Val Arg Gln Ser His Phe Thr Ala Trp Pro Glu His Gly Val
 1025 1030 1035 1040
 Pro Tyr His Ala Thr Gly Leu Leu Ala Phe Ile Arg Arg Val Lys Ala
 1045 1050 1055
 Ser Thr Pro Pro Asp Ala Gly Pro Ile Val Ile His Cys Ser Ala Gly
 1060 1065 1070
 Thr Gly Arg Thr Arg Cys Tyr Ile Val Leu Asp Val Met Leu Asp Met
 1075 1080 1085
 Ala Glu Cys Glu Gly Val Val Asp Ile Tyr Asn Cys Val Lys Thr Leu
 1090 1095 1100
 Cys Ser Arg Arg Val Asn Met Ile Gln Thr Glu Glu Gln Tyr Ile Phe
 1105 1110 1115 1120
 Ile His Asp Ala Ile Leu Glu Ala Cys Leu Cys Gly Glu Thr Thr Ile
 1125 1130 1135
 Pro Val Ser Glu Phe Lys Ala Thr Tyr Lys Glu Met Ile Arg Ile Asp
 1140 1145 1150
 Pro Gln Ser Asn Ser Ser Gln Leu Arg Glu Glu Phe Gln Thr Leu Asn
 1155 1160 1165
 Ser Val Thr Pro Pro Leu Asp Val Glu Glu Cys Ser Ile Ala Leu Leu
 1170 1175 1180
 Pro Arg Asn Arg Asp Lys Asn Arg Ser Met Asp Val Leu Pro Pro Asp
 1185 1190 1195 1200

Arg Cys Leu Pro Phe Leu Ile Ser Thr Asp Gly Asp Ser Asn Asn Tyr
 1205 1210 1215
 Ile Asn Ala Ala Leu Thr Asp Ser Tyr Thr Arg Arg Ser Ala Phe Met
 1220 1225 1230
 Val Thr Leu His Pro Leu Gln Ser Thr Thr Pro Asp Phe Trp Arg Leu
 1235 1240 1245
 Val Tyr Asp Tyr Gly Cys Thr Ser Ile Val Met Leu Asn Gln Leu Asn
 1250 1255 1260
 Gln Ser Asn Ser Ala Trp Pro Cys Leu Gln Tyr Trp Pro Glu Pro Gly
 1265 1270 1275 1280
 Arg Gln Gln Tyr Gly Leu Met Glu Val Glu Phe Met Ser Gly Thr Ala
 1285 1290 1295
 Asp Glu Asp Leu Val Ala Arg Val Phe Arg Val Gln Asn Ile Ser Arg
 1300 1305 1310
 Leu Gln Glu Gly Asp Leu Leu Val Arg His Phe Gln Phe Leu Arg Trp
 1315 1320 1325
 Ser Ala Tyr Arg Asp Thr Pro Asp Ser Lys Lys Ala Phe Leu His Leu
 1330 1335 1340
 Leu Ala Glu Val Asp Lys Trp Gln Ala Glu Ser Gly Asp Gly Arg Thr
 1345 1350 1355 1360
 Ile Val His Cys Leu Asn Gly Gly Arg Ser Gly Thr Phe Cys Ala
 1365 1370 1375
 Cys Ala Thr Val Leu Glu Met Ile Arg Cys His Asn Leu Val Asp Val
 1380 1385 1390
 Phe Phe Ala Ala Gln Thr Leu Arg Asn Tyr Lys Pro Asn Met Val Glu
 1395 1400 1405
 Thr Met Asp Gln Tyr His Phe Cys Tyr Asp Val Ala Leu Glu Tyr Leu
 1410 1415 1420
 Glu Gly Leu Glu Ser Arg
 1425 1430

(2) INFORMATION FOR SEQ ID NO: 35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2810 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 35:

GAATTCCGCA CGAGCGGGCT GGACCTTGCT CGCCCGCGGC GCCATGAGCC GCAGCCTGGA
 60
 CTCGGCGCGG AGCTTCTGAG CGCGGCTGGA AGCGCGGGGC GGCGGGAGGG GGGCAGTCCT
 120
 CGCCGGCGAG TTCAGCGACA TCCAGGCCCTG CTCGGGCGCC TGGAAGGCTG ACCGGCGTGTG
 180

CTCCACCGTG	GCCGGCAGTC	GGCCAGAGAA	CGTGAGGAGA	AACCGCTACA	AAGACGTGCT	240
GCCTTATGAT	CAGACGGAG	TAATCCTCTC	CTCTGTCAG	GAAGGGAC	ACACGGACTA	360
CATTAATGTC	AATTCTTCG	GGGGCGTGG	TGGAAGCTG	GCCTACATTG	CCACGCAGG	420
ACCCCTGCT	CACACCTCG	TAGACTCTG	GAGACTGTC	TGGGAGTTG	GGGTCAGGTT	480
GATCTTGTG	GGCTGTGAG	AGATAGAGAA	TGGGGAGAA	AGGTGTGAGC	GGTACTGGGC	540
CCAGGAGGAG	GAGCCACTG	AGACTGGGT	TTTCTGCA	ACTCTGATAA	AGGAGAAAGTG	600
GCTGAATGAG	GACATCATGC	TCAGGACCC	CAAGGTACA	TTCCAGAAGG	AGTCCCGTTC	660
TGTGTTACCGA	CTACAGTATA	TGTCCTGGCC	AGACCGTGG	GTCCCCAGCA	GTCCCTGACCA	720
CATGCTCGCC	ATGGTGGAGG	AGGCGCTCC	CTTCCAGGG	TCTGGGCTG	AACCCCTCTG	780
TGTCCTACTGC	AGTGGCGGTT	GTGGGCGAAG	AGGCTCTCC	TGACCCGTTG	ATTATGTTGAG	840
GCAGCTGCTC	CTGACCCAGA	TGATCCACCC	TGACTTCTAGT	CTCTTGTAGT	TGGTCTCTAA	900
GATGAGGAAG	CAGCGGCTG	GGCCGTTGCA	GACAGAGAG	CAGTACAGGT	TCTGTACCCA	960
CACGGTGGCT	CAATGTTCT	GCTCCACAT	CCAGAACGTC	AGCCCCCCT	ACCAAGACAT	1020
CAAAAGAAAT	TGTTGGCCAC	TCTACGAGCA	TGCCCCCTC	CTCCGGACTC	CCCAGGCACT	1080
TCTCGCCATA	CCCCCGCCAC	CAGGGAGGG	CCTCAGGAG	ATCTCTGTC	CGGGTCCCC	1140
GGGCCACGCC	ATGGTCGACA	CCTACCGCG	GGAGCAGAAG	CGCCGGGCTC	CAGCGGGGCC	1200
CGGGCGTGGG	ACCGAGAGG	GGACGGGAC	GGGGCGGCG	AGCCCGGAGG	AGGCCCGCC	1260
CTACAGCAAG	GTGACCCAG	GCGCCCGACG	ACCCGGGGC	CACCCGGAGG	ACCGCAGGGG	1320
GACGCTGCTC	GGCCGCGGTC	CTGCTGACCA	AGATCTGCC	GGATCTGGCG	CCTACGAGGA	1380
CGTGGCCGGT	GGAGCTCGAC	CCGGTGGGAC	AGGTTCAC	CTGGCAATTG	GGAGGCCGA	1440
GGGTCCCGCG	GACCCCGCTG	AGTGGTGGAC	CCGGGTGAT	GTCTAACGCC	AGTTCCTCCC	1500
TGTTGCTCTC	TGTTGACTCG	GACTCTGAT	CCCGGGTTC	TGCTGAGCG	CGTGGCAGA	1560
ATGGAAACAG	TGGGGCTGGA	TCAAAGTTAA	AGTTCTCA	GGTGGGAAAT	GTGGGGGCTT	1620
TGCCCCATGA	CTGTCAGCAT	AGACCGCTG	GGCTGGAGG	GGTAGCTAGG	GTATAGTGC	1680
TGGTGAAGGCT	GCACAGAGCA	GATTCAAGAA	AGAAAGAT	GAAGGGGCA	GACCCCTGAG	1740
TTATGAAGGG	GAGAAAGGG	AGATGACCTT	CCGGAGACTC	CTCTCTTCAC	CACACAGCAC	1800
TAGTCATCC	TCAGCACCTG	AGCCTCCCTC	ACTTGGGAC	TCAGGGGAC	ACACAGAGAA	1860
GTGGAATGGAC	ACTTCGGCAT	CCAGGAGCA	CTAACCGAC	CATAACACCA	GCCAAAGCAGA	1920
TTAACCCAG	CGACAGCCGAT	AAAAGAACCT	CCAGATAGGT	AGACAGACAG	ATGGACCAAC	1980
AACTGGACAC	GACAGCCAAA	GCTTCAGAGA	TAACAGTC	AGGTGGACAA	AGGATCCCC	2040
AGCCAGAGAG	AGAGGAGACCA	GCCAAACGCT	TGATAGACCA	GTGCAAGCC	AGAGACCCAC	2100
AAACACAGCC	CCCCAAAGAC	AGACATCTCT	GCTAGCTGG	CAGCGAGGTG	GACCCCTCTAA	2160
GTTAGTCAGA	TTACTAGACA	GATAATAAAG	GATCCCCCTG	TGAAACAGATA	TACAGAGTT	2220
TCAGACCCCA	CTCCCTCAGG	TGGGGCTGGT	GCTGACAGA	CCTCTCTGCC	AGACAGACTC	2280
CTAACCAACCC	AGATGGGATC	CCAGACAGC	AGACATCA	CCACATGGR	TCTTGACATC	2340
CCAGCAGCG	GGCCGAGACTC	TCATCTGAT	GTCTTGATGG	ATGGACCCCA	GCTACTCGAG	2400
CATGATCTG	CGATAGTC	GACAGCTGCC	CCAAATGAGT	ACACATCTCC	AGCTATTCA	2460
ACAGATGGAG	CCCCACGCAA	TCAGGACCTA	TCTAGGGAG	CCCCAGCCAG	ACCCCGGCCA	2520
GACAGACTTC	CAACAGCC	GACCCCTGG	TGTTTCACAC	GCTGCGGAG	TAGCTGGGAC	2580
TACAGGCTCA	ATTTTTTTT	TTTTAAAGAA	ATGAGTTTT	GGCATGTTG	CCAGACTGTT	2640
CTTGAATCTC	CAACCTCGA	AAATCTCC	GCCCTCAGGT	CCCAAATGTC	TGAGATTACA	2700
GGTGTGAGCC	ACCAGGCTCA	GCCCCCTAAC	ATTTGAACAA	CTTTAAATGG	CCCATGGTAG	2810
GGTTCTGCT	AGGATAAAAC	ATTAAGTGGC	TGTTAAAAAGA	AATAAAAGG	GGACACGCT	
CTGTGCAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	AAAAAAAAAA	

(2) INFORMATION FOR SEQ ID NO: 36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 458 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

- (ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 36:

Met Ser Arg Ser Leu Asp Ser Ala Arg Ser Phe Leu Glu Arg Leu Glu
 1 5 10 15
 Ala Arg Gly Arg Glu Gly Ala Val Leu Ala Gly Glu Phe Ser Asp
 20 25 30
 Ile Gln Ala Cys Ser Ala Ala Trp Lys Ala Asp Gly Val Cys Ser Thr
 35 40 45
 Val Ala Gly Ser Arg Pro Glu Asn Val Arg Lys Asn Arg Tyr Lys Asp
 50 55 60
 Val Leu Pro Tyr Asp Gln Thr Arg Val Ile Leu Ser Leu Leu Gln Glu
 65 70 75 80
 Glu Gly His Ser Asp Tyr Ile Asn Gly Asn Phe Ile Arg Gly Val Asp
 85 90 95
 Gly Ser Leu Ala Tyr Ile Ala Thr Gln Gly Pro Leu Pro His Thr Leu
 100 105 110
 Leu Asp Phe Trp Arg Leu Val Trp Glu Phe Gly Val Lys Val Ile Leu
 115 120 125
 Met Ala Cys Arg Glu Ile Glu Asn Gly Arg Lys Arg Cys Glu Arg Tyr
 130 135 140
 Trp Ala Gln Glu Gln Glu Pro Leu Gln Thr Gly Leu Phe Cys Ile Thr
 145 150 155 160
 Leu Ile Lys Glu Lys Trp Leu Asn Glu Asp Ile Met Leu Arg Thr Leu
 165 170 175
 Lys Val Thr Phe Gln Lys Glu Ser Arg Ser Val Tyr Gln Leu Gln Tyr
 180 185 190
 Met Ser Trp Pro Asp Arg Gly Val Pro Ser Ser Pro Asp His Met Leu
 195 200 205
 Ala Met Val Glu Glu Ala Arg Arg Leu Gln Gly Ser Gly Pro Glu Pro
 210 215 220
 Leu Cys Val His Cys Ser Ala Gly Cys Gly Arg Thr Gly Val Leu Cys
 225 230 235 240
 Thr Val Asp Tyr Val Arg Gln Leu Leu Leu Thr Gln Met Ile Pro Pro
 245 250 255
 Asp Phe Ser Leu Phe Asp Val Val Leu Lys Met Arg Lys Gln Arg Pro
 260 265 270
 Ala Ala Val Gln Thr Glu Glu Gln Tyr Arg Phe Leu Tyr His Thr Val
 275 280 285
 Ala Gln Met Phe Cys Ser Thr Leu Gln Asn Ala Ser Pro His Tyr Gln
 290 295 300
 Asn Ile Lys Glu Asn Cys Ala Pro Leu Tyr Asp Asp Ala Leu Phe Leu
 305 310 315 320

Arg Thr Pro Gln Ala Leu Leu Ala Ile Pro Arg Pro Pro Gly Gly Val
 325 330 335
 Leu Arg Ser Ile Ser Val Pro Gly Ser Pro Gly His Ala Met Ala Asp
 340 345 350
 Thr Tyr Ala Glu Glu Gln Lys Arg Gly Ala Pro Ala Gly Ala Gly Ser
 355 360 365
 Gly Thr Gln Thr Gly Thr Gly Ala Arg Ser Ala Glu Glu Ala
 370 375 380
 Pro Leu Tyr Ser Lys Val Thr Pro Arg Ala Gln Arg Pro Gly Ala His
 385 390 395 400
 Ala Glu Asp Ala Arg Gly Thr Leu Pro Gly Arg Val Pro Ala Asp Gln
 405 410 415
 Ser Pro Ala Gly Ser Gly Ala Tyr Glu Asp Val Ala Gly Gly Ala Gln
 420 425 430
 Thr Gly Gly Leu Gly Phe Asn Leu Arg Ile Gly Arg Pro Lys Gly Pro
 435 440 445
 Arg Asp Pro Pro Ala Glu Trp Thr Arg Val
 450 455

(2) INFORMATION FOR SEQ ID NO: 37:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 503 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 37:

Met Glu Pro Ala Gly Pro Ala Pro Gly Arg Leu Gly Pro Leu Leu Cys
 1 5 10 15
 Leu Leu Leu Ala Ala Ser Cys Ala Trp Ser Gly Val Ala Gly Glu Glu
 20 25 30
 Glu Leu Gln Val Ile Gln Pro Asp Lys Ser Val Ser Val Ala Ala Gly
 35 40 45
 Glu Ser Ala Ile Leu His Cys Thr Val Thr Ser Leu Ile Pro Val Gly
 50 55 60
 Pro Ile Gln Trp Phe Arg Gly Ala Gly Pro Ala Arg Glu Leu Ile Tyr
 65 70 75 80
 Asn Gln Lys Glu Gly His Phe Pro Arg Val Thr Thr Val Ser Glu Ser
 85 90 95
 Thr Lys Arg Glu Asn Met Asp Phe Ser Ile Ser Ile Ser Asn Ile Thr
 100 105 110

Pro Ala Asp Ala Gly Thr Tyr Cys Val Lys Phe Arg Lys Gly Ser
 115 120 125
 Pro Asp Thr Glu Phe Lys Ser Gly Ala Gly Thr Glu Leu Ser Val Arg
 130 135 140
 Ala Lys Pro Ser Ala Pro Val Val Ser Gly Pro Ala Ala Arg Ala Thr
 145 150 155 160
 Pro Gln His Thr Val Ser Phe Thr Cys Glu Ser His Gly Phe Ser Pro
 165 170 175
 Arg Asp Ile Thr Leu Lys Trp Phe Lys Asn Gly Asn Glu Leu Ser Asp
 180 185 190
 Phe Gln Thr Asn Val Asp Pro Val Gly Glu Ser Val Ser Tyr Ser Ile
 195 200 205
 His Ser Thr Ala Lys Val Val Leu Thr Arg Glu Asp Val His Ser Gln
 210 215 220
 Val Ile Cys Glu Val Ala His Val Thr Leu Gln Gly Asp Pro Leu Arg
 225 230 235 240
 Gly Thr Ala Asn Leu Ser Glu Thr Ile Arg Val Pro Pro Thr Leu Glu
 245 250 255
 Val Thr Gln Gln Pro Val Arg Ala Glu Asn Gln Val Asn Val Thr Cys
 260 265 270
 Gln Val Arg Lys Phe Tyr Pro Gln Arg Leu Gln Leu Thr Trp Leu Glu
 275 280 285
 Asn Gly Asn Val Ser Arg Thr Glu Thr Ala Ser Thr Val Thr Glu Asn
 290 295 300
 Lys Asp Gly Thr Tyr Asn Trp Met Ser Trp Leu Leu Val Asn Val Ser
 305 310 315 320
 Ala His Arg Asp Asp Val Lys Leu Thr Cys Gln Val Glu His Asp Gly
 325 330 335
 Gln Pro Ala Val Ser Lys Ser His Asp Leu Lys Val Ser Ala His Pro
 340 345 350
 Lys Glu Gln Gly Ser Asn Thr Ala Ala Glu Asn Thr Gly Ser Asn Glu
 355 360 365
 Arg Asn Ile Tyr Ile Val Val Gly Val Val Cys Thr Leu Leu Val Ala
 370 375 380
 Leu Leu Met Ala Ala Leu Tyr Leu Val Arg Ile Arg Gln Lys Lys Ala
 385 390 395 400
 Gln Gly Ser Thr Ser Ser Thr Arg Leu His Glu Pro Glu Lys Asn Ala
 405 410 415
 Arg Glu Ile Thr Gln Asp Thr Asn Asp Ile Thr Tyr Ala Asp Leu Asn
 420 425 430

Leu Pro Lys Gly Lys Lys Pro Ala Pro Gln Ala Ala Glu Pro Asn Asn
 435 440 445
 His Thr Glu Tyr Ala Ser Ile Gln Thr Ser Pro Gln Pro Ala Ser Glu
 450 455 460
 Asp Thr Leu Thr Tyr Ala Asp Leu Asp Met Val His Leu Asn Arg Thr
 465 470 475 480
 Pro Lys Gln Pro Ala Pro Lys Pro Glu Pro Ser Phe Ser Glu Tyr Ala
 485 490 495
 Ser Val Gln Val Pro Arg Lys
 500

(2) INFORMATION FOR SEQ ID NO: 38:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 398 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(xi) SEQUENCE DESCRIPTION: SEQ ID NO: 38:

Met Pro Val Pro Ala Ser Trp Pro His Leu Pro Ser Pro Phe Leu Leu
 1 5 10 15
 Met Thr Leu Leu Gly Arg Leu Thr Gly Val Ala Gly Glu Asp Glu
 20 25 30
 Leu Gln Val Ile Gln Pro Glu Lys Ser Val Ser Val Ala Ala Gly Glu
 35 40 45
 Ser Ala Thr Leu Arg Cys Ala Met Thr Ser Leu Ile Pro Val Gly Pro
 50 55 60
 Ile Met Trp Phe Arg Gly Ala Gly Arg Glu Leu Ile Tyr Asn
 65 70 75 80
 Gln Lys Glu Gly His Phe Pro Arg Val Thr Thr Val Ser Glu Leu Thr
 85 90 95
 Lys Arg Asn Asn Leu Asn Phe Ser Ile Ser Ile Ser Asn Ile Thr Pro
 100 105 110
 Ala Asp Ala Gly Thr Tyr Tyr Cys Val Lys Phe Arg Lys Gly Ser Pro
 115 120 125
 Asp Asp Val Glu Phe Lys Ser Gly Ala Gly Thr Glu Leu Ser Val Arg
 130 135 140
 Ala Lys Pro Ser Ala Pro Val Val Ser Gly Pro Ala Val Arg Ala Thr
 145 150 155 160
 Pro Glu His Thr Val Ser Phe Thr Cys Glu Ser His Gly Phe Ser Pro
 165 170 175

Arg Asp Ile Thr Leu Lys Trp Phe Lys Asn Gly Asn Glu Leu Ser Asp
180 185 190

Phe Gln Thr Asn Val Asp Pro Ala Gly Asp Ser Val Ser Tyr Ser Ile
195 200 205

His Ser Thr Ala Arg Val Val Leu Thr Arg Gly Asp Val His Ser Gln
210 215 220

Val Ile Cys Glu Met Ala His Ile Thr Leu Gln Gly Asp Pro Leu Arg
225 230 235 240

Gly Thr Ala Asn Leu Ser Glu Ala Ile Arg Val Pro Pro Thr Leu Glu
245 250 255

Val Thr Gln Gln Pro Met Arg Ala Glu Asn Gln Ala Asn Val Thr Cys
260 265 270

Gln Val Ser Asn Phe Tyr Pro Arg Gly Leu Gln Leu Thr Trp Leu Glu
275 280 285

Asn Gly Asn Val Ser Arg Thr Glu Thr Ala Ser Thr Leu Thr Glu Asn
290 295 300

Lys Asp Gly Thr Tyr Asn Trp Met Ser Trp Leu Leu Val Asn Thr Cys
305 310 315 320

Ala His Arg Asp Asp Val Val Leu Thr Cys Gln Val Glu His Asp Gly
325 330 335

Gln Gln Ala Val Ser Lys Ser Tyr Ala Leu Glu Ile Ser Ala His Gln
340 345 350

Lys Glu His Gly Ser Asp Ile Thr His Glu Pro Ala Leu Ala Pro Thr
355 360 365

Ala Pro Leu Leu Val Ala Leu Leu Leu Gly Pro Lys Leu Leu Leu Val
370 375 380

Val Gly Val Ser Ala Ile Tyr Ile Cys Trp Lys Gln Lys Ala
385 390 395